Comparison of some Different Methods for Hypothesis Test of Means of Log-normal Populations

Saba Aghadoust, Kamel Abdollahnezhad*, Farhad Yaghmaei, Ali Akbar Jafari
Department of Statistics, Golestan University, Gorgan, Iran
Department of Statistics, Yazd University, Yazd, Iran

Abstract

The log-normal distribution is used to describe the positive data, that it has skewed distribution with small mean and large variance. This distribution has application in many sciences for example medicine, economics, biology and alimentary science, ect. Comparison of means of several log-normal populations always has been in focus of researchers, but the test statistic are not easy to derive or extremely complicated for this comparisons. In this paper, the different methods exist for this testing that we can point out F-test, likelihood ratio test, generalized p-value approach and computational approach test. In this line with help of simulation studies, in this methods we compare and evaluate size and power test.

Keywords: Log-normal distribution, Hypothesis test, Size of a test, Power of a test, Maximum likelihood estimation.
2010 AMS Subject Classification: 62F03

1 Introduction

The skewed distributions are particularly common when mean values are small, variances are large and values cannot be negative (for example lengths of latent periods of infectious diseases), and often closely fit the log-normal distribution. The log-normal distribution has been widely used in medical, biological and economic studies, where data are positive and have a right-skewed distribution

Let X_{ij} be random sample from k independent log-normal distributions, i.e.

$$Y_{ij} = \log(X_{ij}) \ \sim N\left(\mu_i, \sigma_i^2\right), \quad i = 1, \ldots, k;; \ j = 1, \ldots, n_i.$$

*Corresponding: k.abdollahnezhad@gu.ac.ir

Also, let $\varphi_i = E(X_{ij}) = \exp(\mu_i + \frac{1}{2}\sigma_i^2)$ denote the mean of the i-th population. Suppose we are interested in testing

$$H_0: \ \varphi_1 = \varphi_2 = \ldots = \varphi_k \qquad \text{vs.} \qquad H_A: \text{Not all the } \varphi_i s \text{ are equal.} \tag{1.1}$$

Then the testing problem (1.1) is equivalent to testing

$$H_0: \eta_i = \eta, \qquad \text{vs.} \qquad H_A\text{: Not all the } \eta_i s \text{ are equal} \tag{1.2}$$

where $\eta_i = \log(\varphi_i) \ = \mu_i + \frac{1}{2}\sigma_i^2$, and η is unspecified.

It is well-known that the maximum likelihood estimators (MLE's) for μ_i maximum likelihood estimation and σ_i^2 are $\overline{Y}_i$ and S_i^2, respectively, where

$$\overline{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} \qquad \text{and} \qquad S_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_i\right)^2, \tag{1.3}$$

Therefore, the MLE of η_i is $\widehat{\eta}_i = \widehat{\mu}_i + \frac{1}{2}\widehat{\sigma}_i^2$ that it has approximately normal distribution with mean η_i and variance $v_i = \frac{\sigma_i^2}{n_i} + \frac{(n_i-1)\sigma_i^4}{2n_i^2}$.

2 The CAT

Pal et al. (2007) introduced the CAT in a general setup. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a population with the known density function $f(x|\lambda)$ with $\lambda = (\theta, \psi)$, where θ is the parameter of interest and ψ is the nuisance parameter. To test $H_0 \ : \theta = \theta_0$ against a suitable alternative H_1 at level a, the general methodology of the CAT for testing is given through the following steps:

1. First obtain $\widehat{\lambda} = \left(\hat{\theta}, \widehat{\psi}\right)$, the MLE of λ.

2. Assume that H_0 is true, i.e., set $\theta = \theta_0$. Then find the MLE of ψ from the data. Call this as the 'restricted MLE' of ψ under H_0 and denote by $\widehat{\psi}_{RML}$.

3. Generate artificial sample $Y_1, Y_2, \ldots, Y_n$ from $f(x|\theta_0, \widehat{\psi}_{RML})$ a large number of times (say, M times). For each of these replicated samples, recalculate the MLE of $\lambda = (\theta, \psi)$. Retain only the component that is relevant for .Let these recalculated MLE values of be $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_M$.

4. Let $\hat{\theta}_{(1)} < \hat{\theta}_{(2)} < \cdots < \hat{\theta}_{(M)}$ be the ordered values of $\hat{\theta}_l, \ 1 \le l \le M$.

5. (i) For testing $H_0 \; : \theta = \theta_0$ against $H_1 \; : \theta < \theta_0$, define $\widehat{\theta}_L = \widehat{\theta}_{(\alpha M)}$. Reject H_0 if $\hat{\theta} < \widehat{\theta}_L$ and accept H_0 otherwise. Alternatively, calculate the p-value as

$$p = \frac{1}{M} \sum_{l=1}^{M} I(\hat{\theta}_l < \hat{\theta}).$$

(ii) For testing $H_0 \; : \theta = \theta_0$ against $H_1 \; : \theta > \theta_0$, define $\widehat{\theta}_U = \widehat{\theta}_{((1-\alpha)M)}$. Reject H_0 if $\hat{\theta} > \widehat{\theta}_U$ and accept H_0 otherwise. Alternatively, calculate the p-value as

$$p = \frac{1}{M} \sum_{l=1}^{M} I(\hat{\theta}_l > \hat{\theta}).$$

(iii) For testing $H_0 \; : \theta = \theta_0$ against $H_1 \; : \theta < \theta_0$, define $\widehat{\theta}_L = \widehat{\theta}_{((\alpha/2)M)}$ and $\widehat{\theta}_U = \widehat{\theta}_{((1-\alpha/2)M)}$. Reject H_0 if $\hat{\theta} < \widehat{\theta}_L$ or $\hat{\theta} > \widehat{\theta}_U$ and accept H_0 otherwise. Alternatively, calculate the p-value as:

$$p = 2\min\left(p_1, 1 - p_1\right) \; ,$$

where $p_1 = \frac{1}{M} \sum_{l=1}^{M} I(\hat{\theta}_l < \hat{\theta})$.

Now implement the CAT for testing the equality of several log-normal means. To apply our proposed CAT, we first need to express H_0 in term of a suitable scalar θ. Define

$$\theta = h\left(\mu_i; ; \sigma_i^2\right) = \sum_{i=1}^{k} \frac{1}{v_i} (\eta_i - \overline{\eta})^2 = \sum_{i=1}^{k} \frac{\eta_i^2}{v_i} - \frac{\left(\sum_{i=1}^{k} \frac{\eta_i}{v_i}\right)^2}{\sum_{i=1}^{k} \frac{1}{v_i}},$$

where $\overline{\eta} = \left(\sum_{i=1}^{k} \frac{1}{v_i}\right)^{-1} \sum_{i=1}^{k} \frac{\eta_i}{v_i}$. It is seen that the hypothesis in (1.2) is equivalent to

$$H_0^* : \theta = 0 \qquad vs \quad H_A^* : \theta > 0.$$

If we apply the steps of CAT, then we have the following steps for testing the equality of means of several log-normal distributions:

1) Obtain $\widehat{\mu}_i = \overline{X}_i$and $\widehat{\sigma}_i^2 = S_i^2$, $i = 1, \ldots, k$, and calculate $\widehat{\theta} = h\left(\widehat{\mu}_i; ; \widehat{\sigma}_i^2\right)$.

2) Assume that H_0 in (1.2) is true, i.e., set $\mu_i = \eta - \frac{1}{2}\sigma_i^2, \; 1 = i = k$. The likelihood function (under H_0) is a function of $\left(\eta, \sigma_1^2, \ldots, \sigma_k^2\right)$ only. The restricted MLE's of $\eta, \sigma_1^2, \ldots, \sigma_k^2$, denoted by $\widehat{\eta}_{RML}, \widehat{\sigma}^2_{i(RML)}, \; 1 = i = k$, are found using numerical methods, see [6]. Define $\widehat{\mu}_{i(RML)} = \widehat{\eta}_{RML} - \frac{1}{2}\widehat{\sigma}^2_{i(RML)}$

3) Generate artificial sample $Y_{i1}, \ldots, Y_{in_i} \, (= \mathbf{Y_i}, say)$ independent random sample from $N\left(\widehat{\mu}_{i(RML)}, \widehat{\sigma}^2_{i(RML)}\right)$. Repeat this process M times. In the l-th replication $(1 = l = M)$ based on $Y_i^{(l)}$ get the MLE's of μ_i and σ_i^2 by (1.3) and call them as $\widehat{\mu}_{0i}^{(l)}$ and $\widehat{\sigma}_{0i}^{2(l)}$. Then recalculate $\widehat{\theta}$ as $\widehat{\theta}_{0l} = h(\widehat{\mu}_{0i}^{(l)}, \widehat{\sigma}_{0i}^{2(l)})$.

4) Order the $\widehat{\theta}_{0l}$ values as $\widehat{\theta}_{0(1)} = \widehat{\theta}_{0(2)} = . . . = \widehat{\theta}_{0(M)}$.

5) Let $\widehat{\theta}_U = \widehat{\theta}_{0((1-\alpha)M)}$ and reject H_0 if $\hat{\theta} > \widehat{\theta}_U$ and accept H_0 otherwise. Alternatively, if the value $p = \frac{1}{M}\sum_{l=1}^{M} I\left(\widehat{\theta}_{0l} > \widehat{\theta}_{ML}\right)$ is smaller than the nominal level a, then reject H_0.

References

[1] Abdollahnezhad, K., Babanezhad, M., Jafari A. A. (2012). Inference on difference of means of two lognormal distributions a generalized approach.*Journal of Statisical and Econemtric Methods*, **1**, 125-131

[2] Behboodian, J. and Jafari, A. A. (2006). Generalized inference for the common mean of several lognormal populations. *Journal of Statistical Theory and Applications*, **5(3)**, 240-259.

[3] Chang, C. and Pal, N. (2008). A revisit to the Behrens-Fisher problem: comparison of five test methods, *Communications in Statistics-Simulation and Computation* **37**, 1064-1085.

[4] Chang, C., Pal, N., Lim, W. K. and Lin, J. J. (2010). Comparing several population means: a parametric bootstrap method , and its comparison with usual ANOVA F test as well as ANOM, *Computational Statistics* **25(1)**, 71-95.

[5] Gill, P. S. (2004). Small-sample inference for the comparison of means of log-normal distributions, *Biometrics* **60**, 525-527.

[6] Jafari, A. A. (2012). Inferences on the ratio of two generalized variances: independent and correlated cases. *Statistical Methods and Applications*, **21(3)**, 297-314.

[7] Jafari, A. A., Abdollahnezhad, K. (2014). Inferences on the means of two log-normal distributions; a computational approach test. *Communications in Statistics- Simulation and Computation*. Accepted manuscript.

[8] Krishnamoorthy, K. and Mathew, T. (2003), Inference on the means of log-normal distributions using generalized P-value and generalized confidence intervals , *Journal of Statistical Planning and Infernce* **115**, 103-121.

[9] Li, X. (2009), A generalized p-value approach for comparing the means of several log-normal populations, *Statistics and Probability Letters* **79**, 1404-1408.

[10] Lin, S. H., Lee, J. C. and Wang, R. S. (2007), Generalized inferences on the common mean vector of several multivariate normal populations, *Journal of Statistical Planning and Infernce* **137**, 2240-2249.

[11] Lin, S. H. and Wang, R. S. (2009), Inferences on a linear combination of k multivariate normal mean vector, *J. Appl. Statist.* **36**, 415-428.

[12] Lin, S. H. and Wang, R. S. (2013), Modified method on the means for several log-normal distributions, *Biometrics* **60**, 525-527.

[13] Pal, N., Lim, W. K. and Ling, C. H. (2007), A Computational approach to statistical inferences, *Journal of Applied Probability and Statistics*, **2(1)**, 13-35.

[14] Scovgaard, I. M. (2001), Likelihood asymptotics, *Scandinavian Journal of Statistics* **28**, 3-32.

[15] Soltani, A. R. and Abdollahnezhad, K. (2013). Testing the log-normal mean: comparison of four test methods, *Journal of Applied Probability and Statistics*, **8(1)**, 1-10.

[16] Tian, L. and Wu, J. (2007), Inferences on the common mean of several multivariate log-normal populations: the generalized variable approach, *Biometrical Journal*, **49**, 944-951.

[17] Tsui, K. W. and Weerahadi, S. (1989), Generalized p-value in significance testing of hypothesis in the presence of nuisance parameters, *Journal of the American Statistical Association*, **84**, 602-607.

[18] Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association*, **88(423)**, 899-905.

[19] Zhou, X. H., Gao, S. and Hui, S. L. (1997). Methods for comparing the means of two independent lognormal samples, *Biometrics*, **53**, 1129-1135.